Improving the readability of clustered social networks using node duplication

Nathalie Henry,
Anastasia Bezerianos,
& Jean-Daniel Fekete

The University of Sydney
Australia

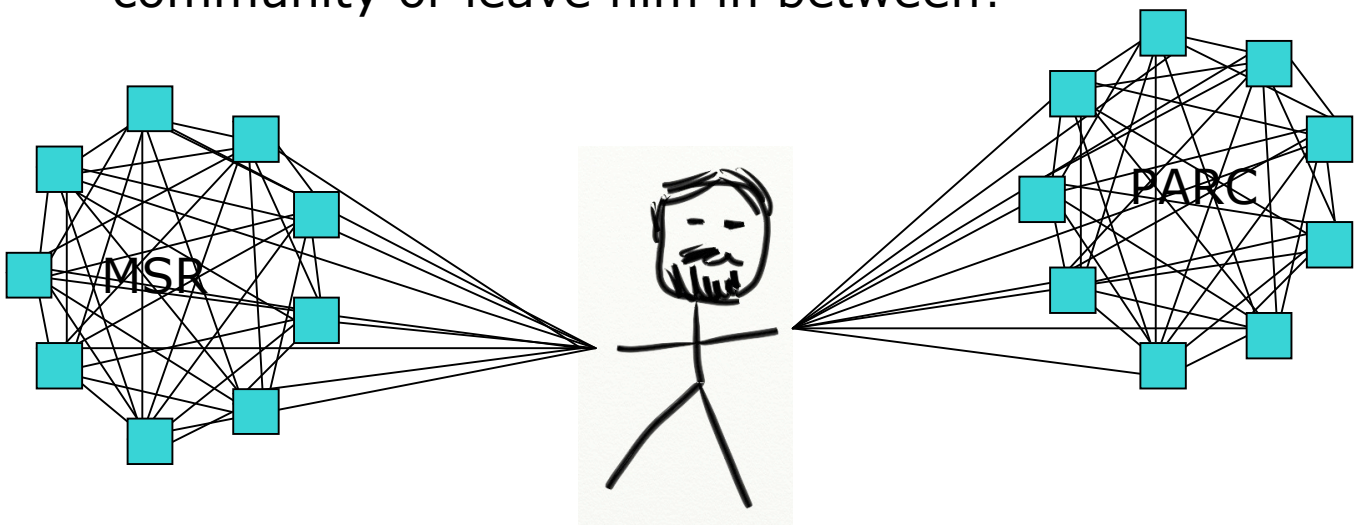
Nathalie.Henry@lri.fr

The Problem

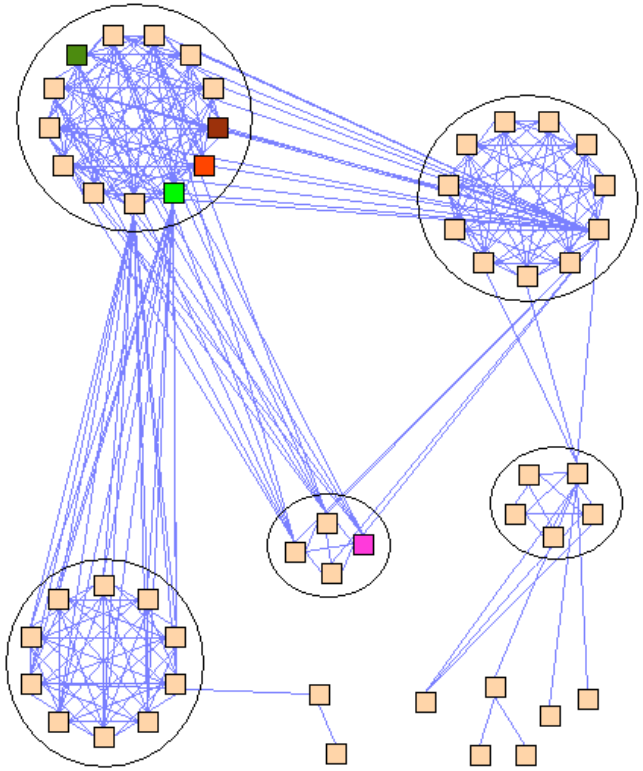
Analyzing Social Networks

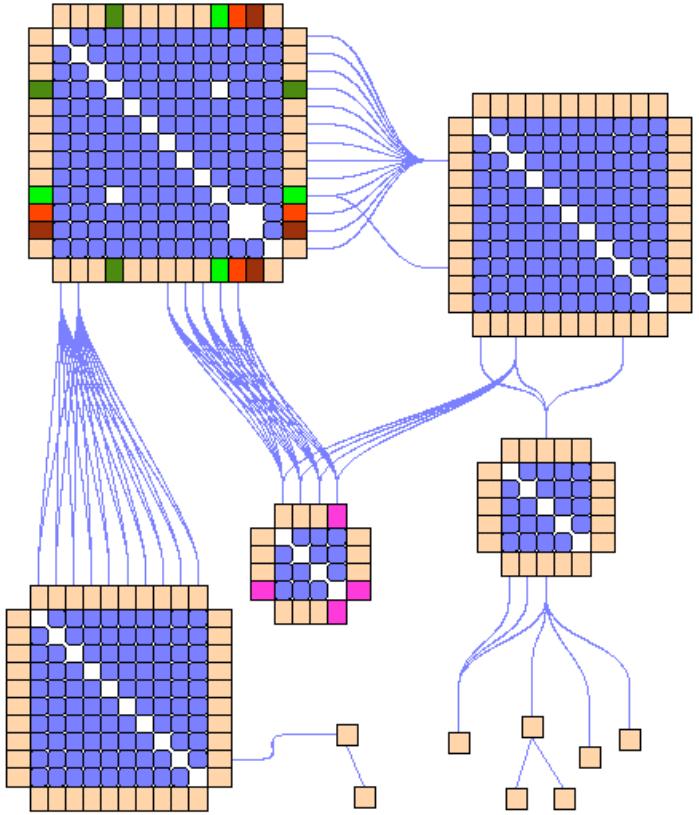
- What are we going to do with George?

Placing him in one or the other
community or leave him in between?



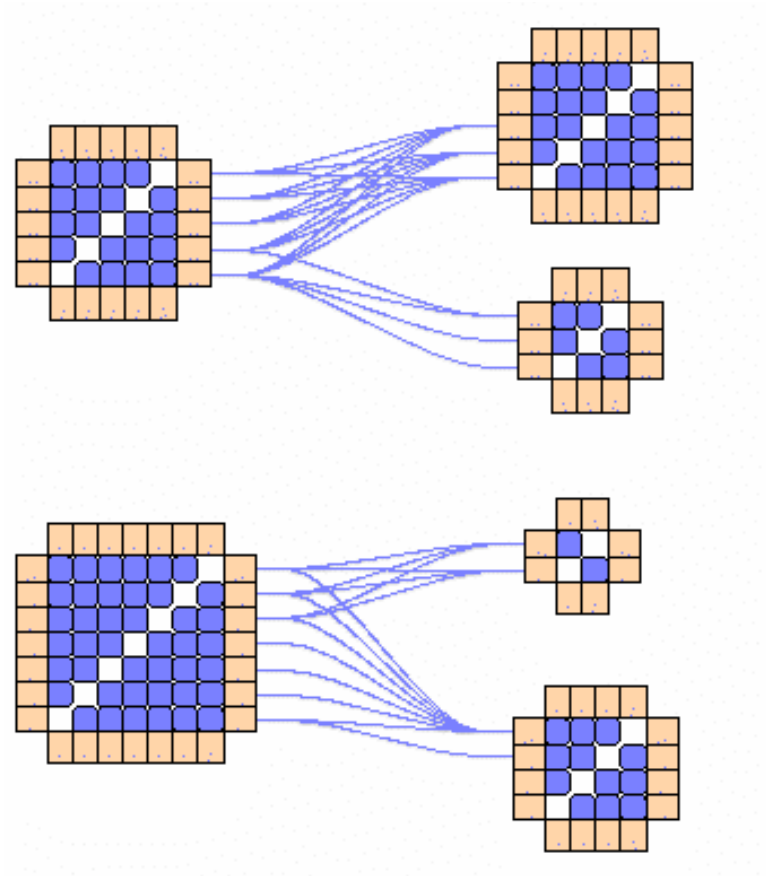
Improving intra-community readability

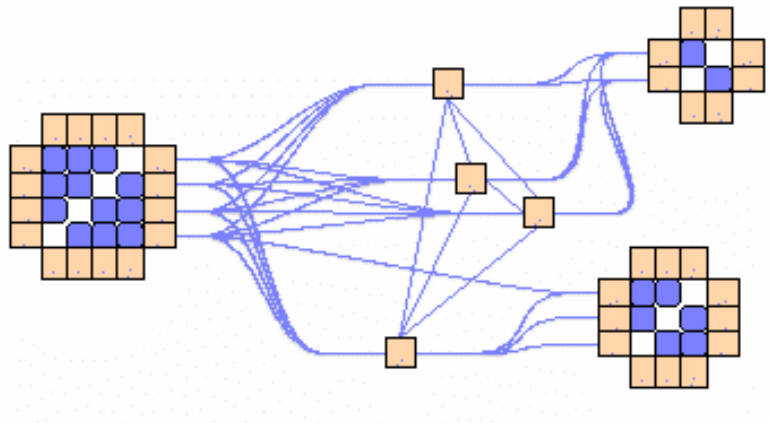




[Henry, Fekete, McGuffin 07]

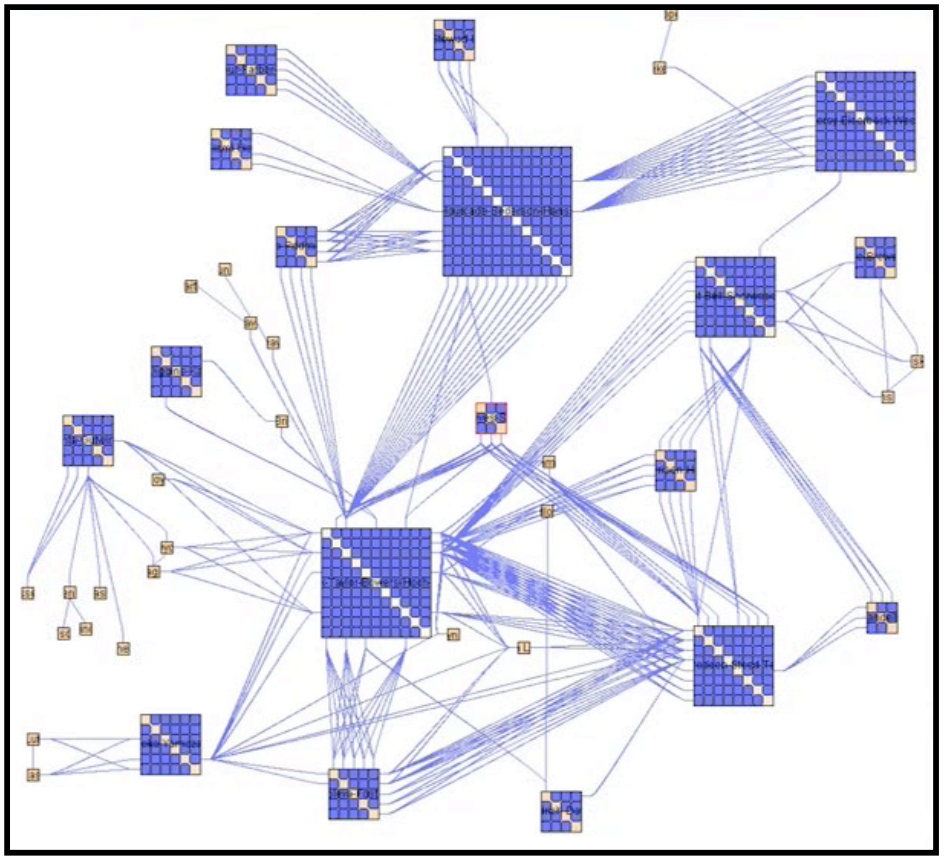
Impact on community perception



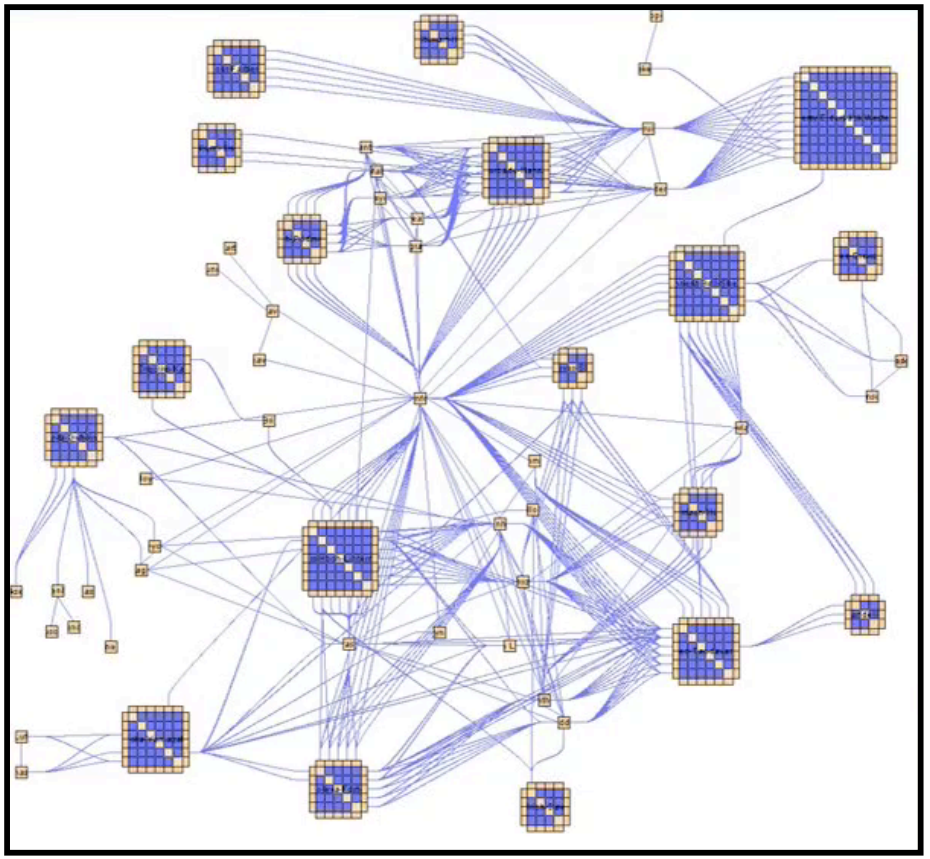


Same portion of network with central actors placed in one or the other community or extracted in between.

Impact on readability



Central actors in one community

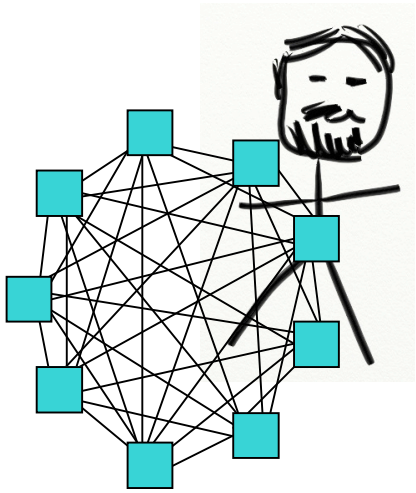


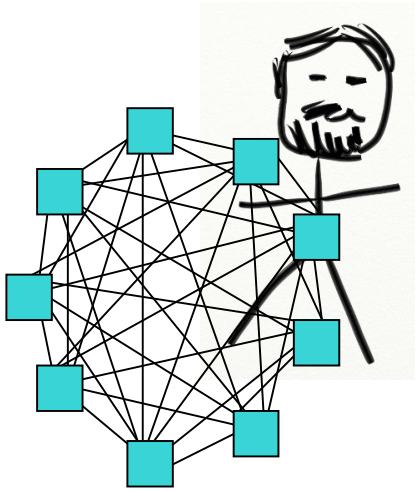
Central actors extracted

Our Solution

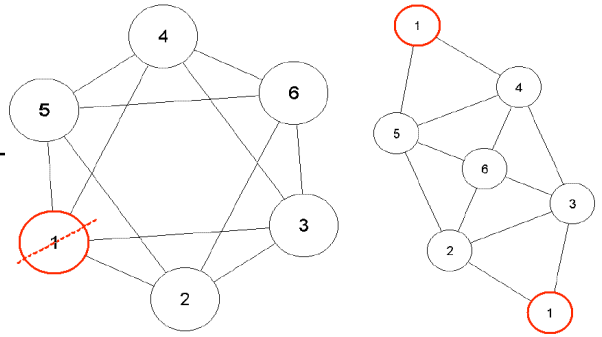
The Solution

- Let’s clone George!



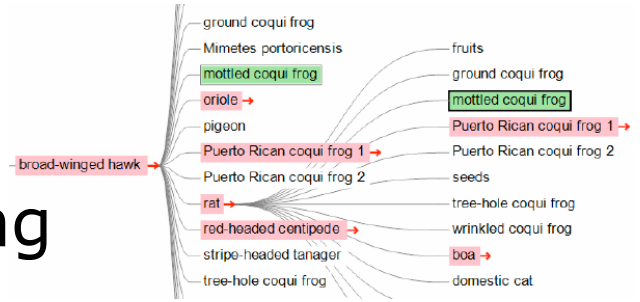


Duplication in the litterature

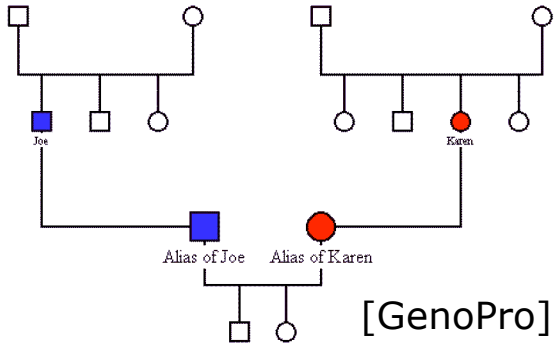


[Eades, Mendonça 95]

- Duplications as dataset errors
- Improving readability
  - Vertex splitting in graph drawing
  - Clones in TreePlus or Ontorama
  - Alias for genealogical tools

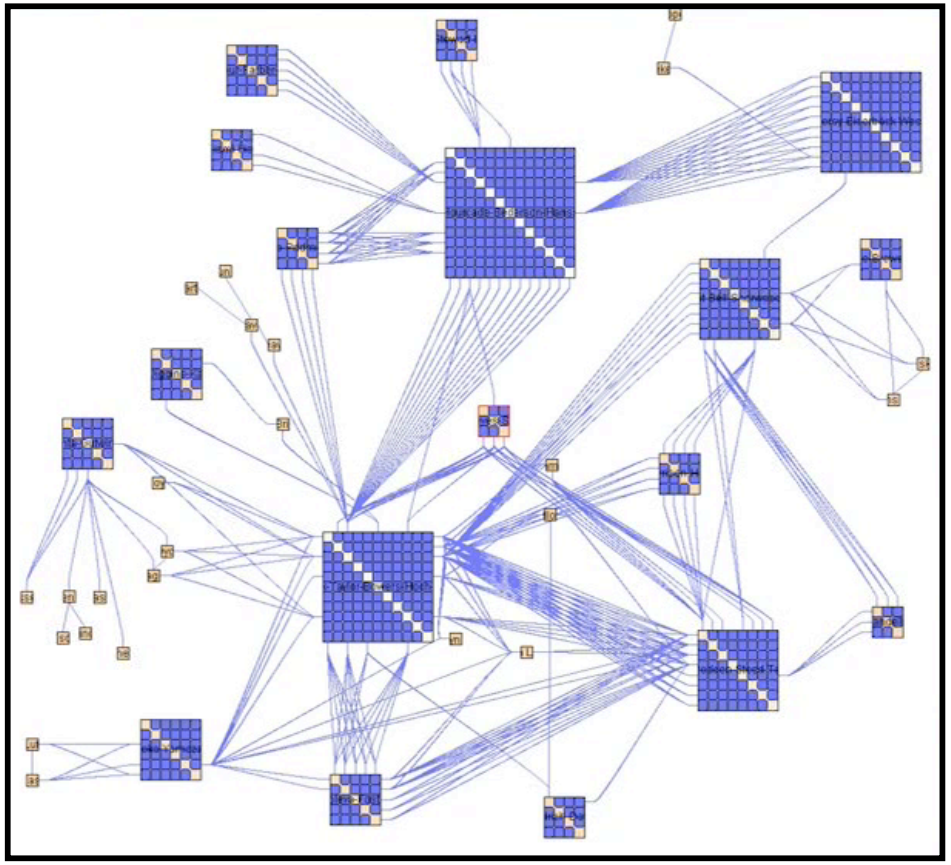


[Lee et al. 06]

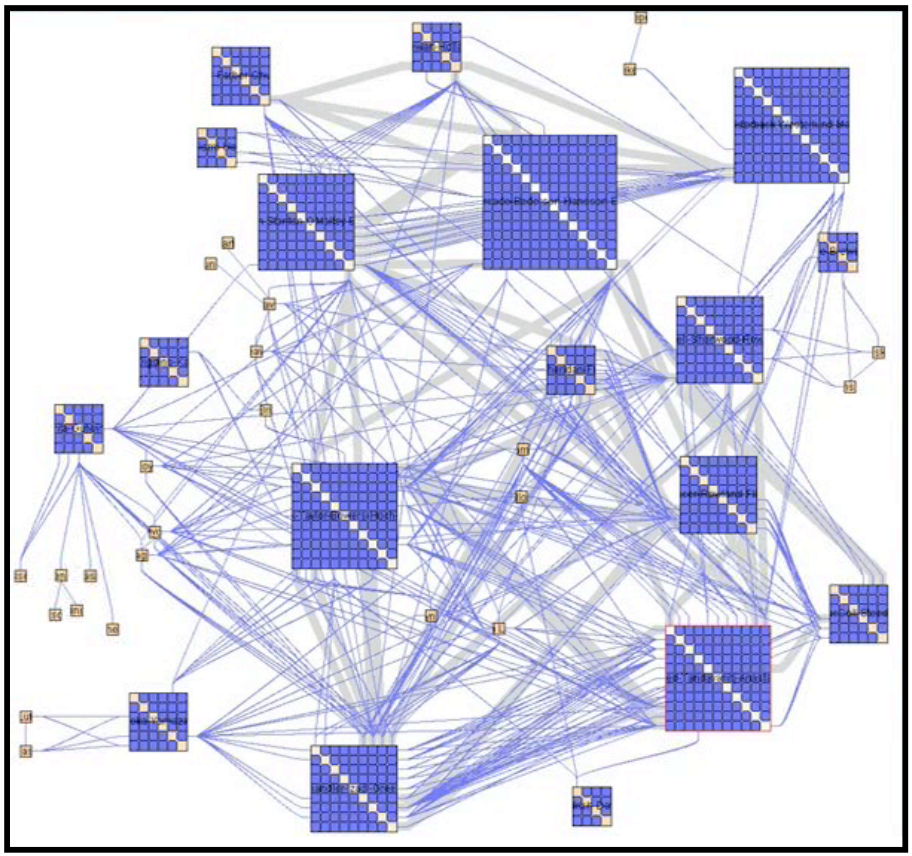


[GenoPro]

Introducing duplicates in a clustered graph



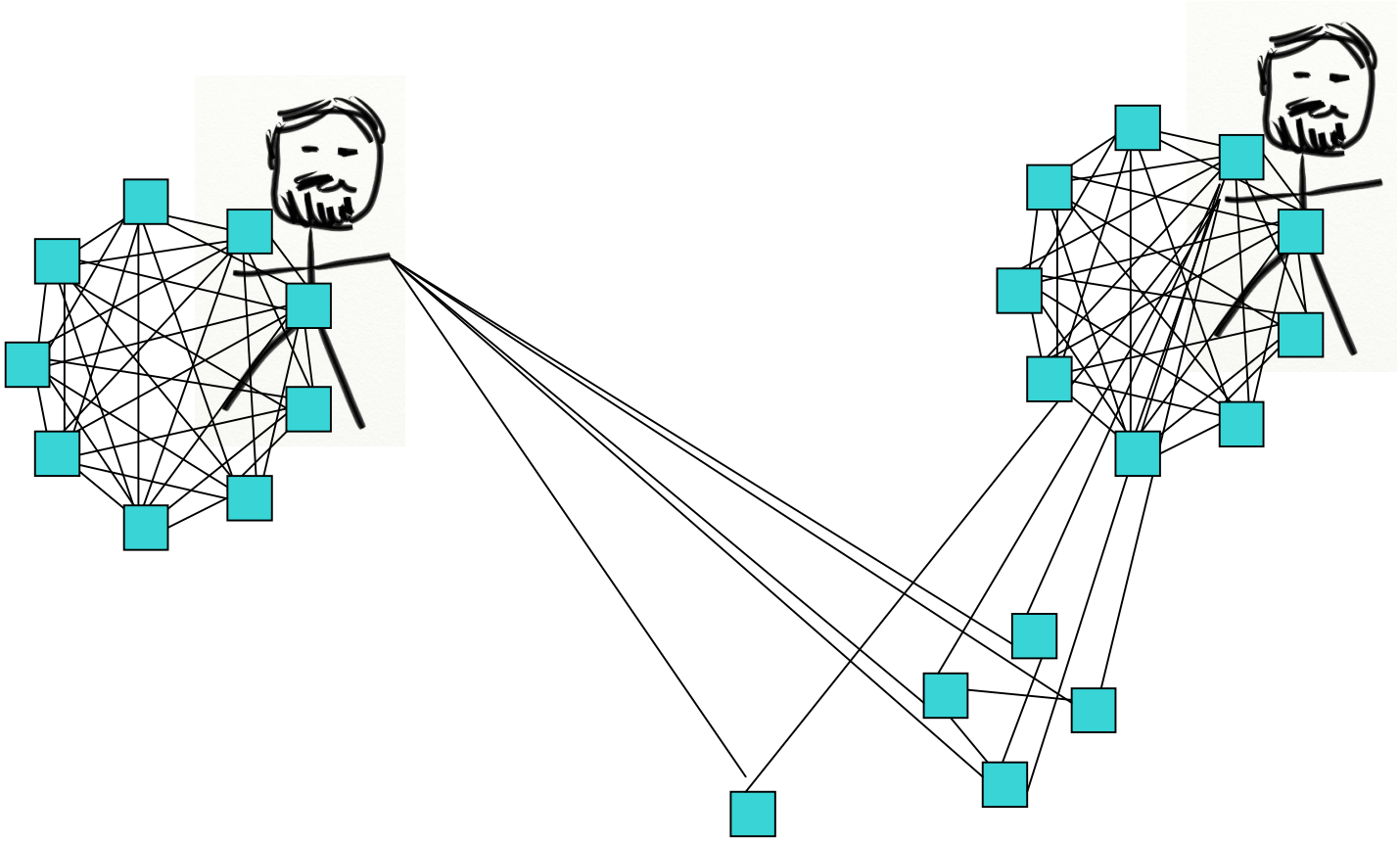
Central actors in one community



Central actors duplicated

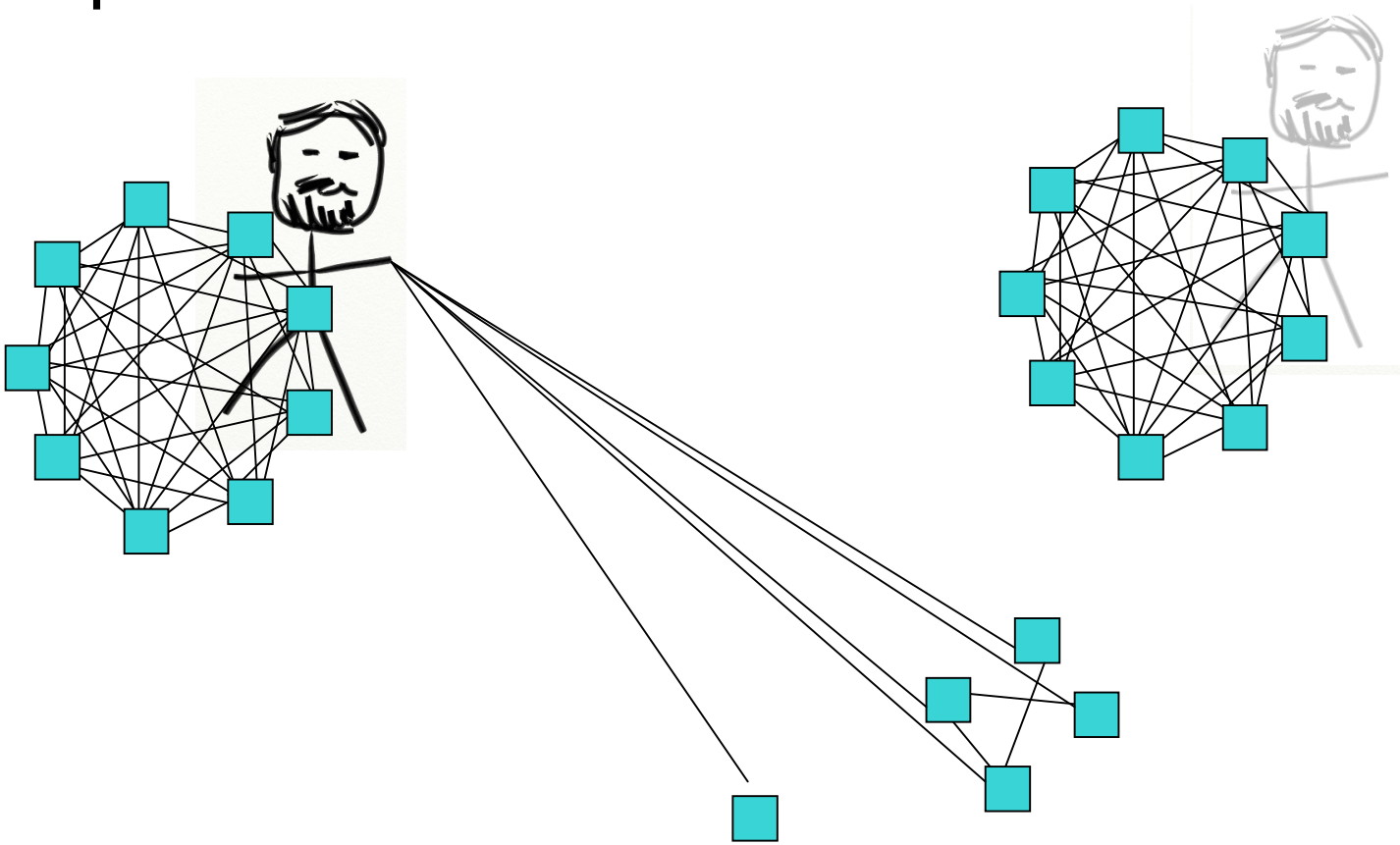
Two types of duplications

- Clone duplication



Two types of duplications

- Split duplication



Visual Representation

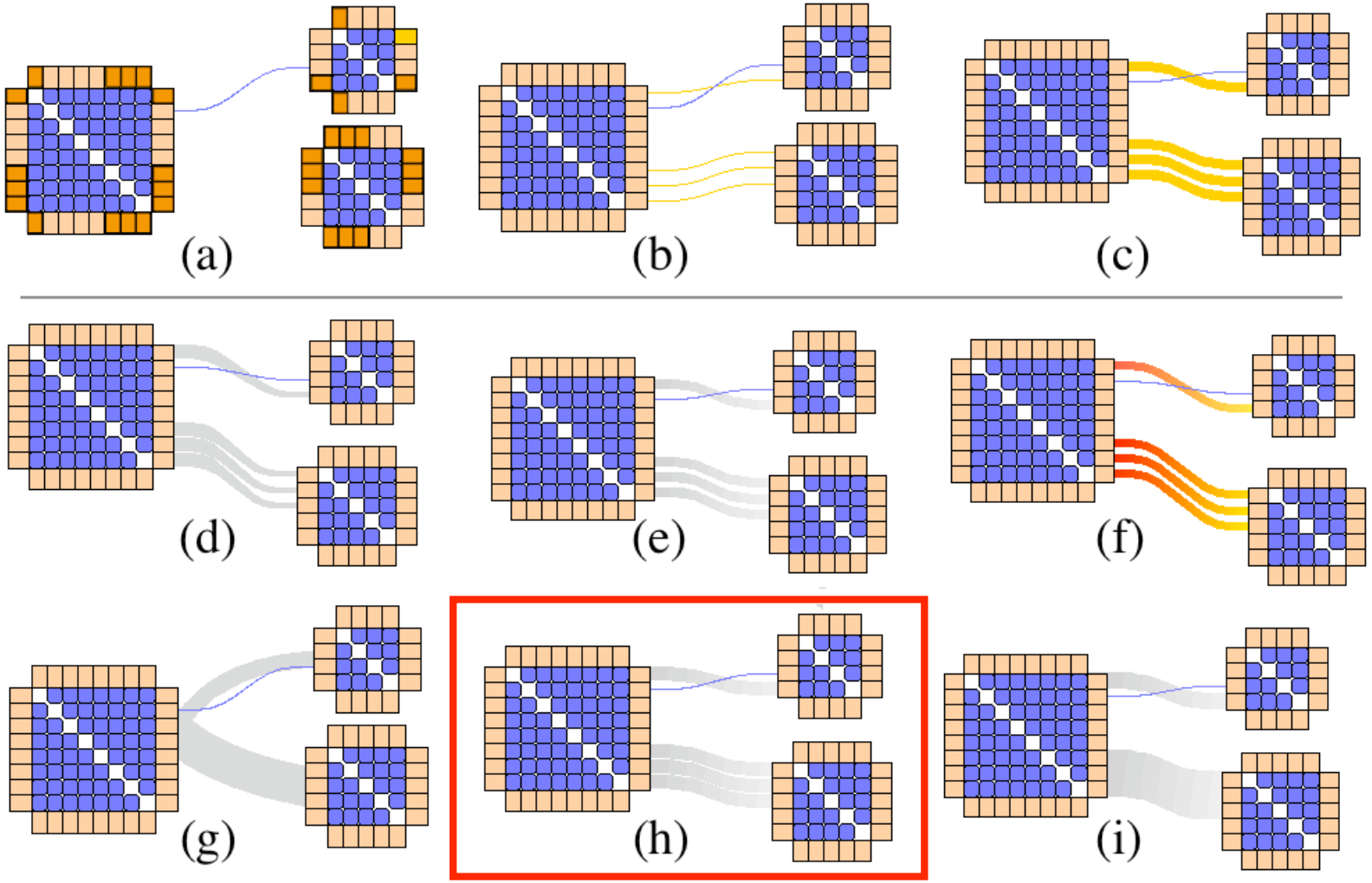
- In the litterature
  - Node coloring
  - Links between duplicates

- Our design goals
  - Maximize the awareness of duplicates
  - Minimize interference with regular network links

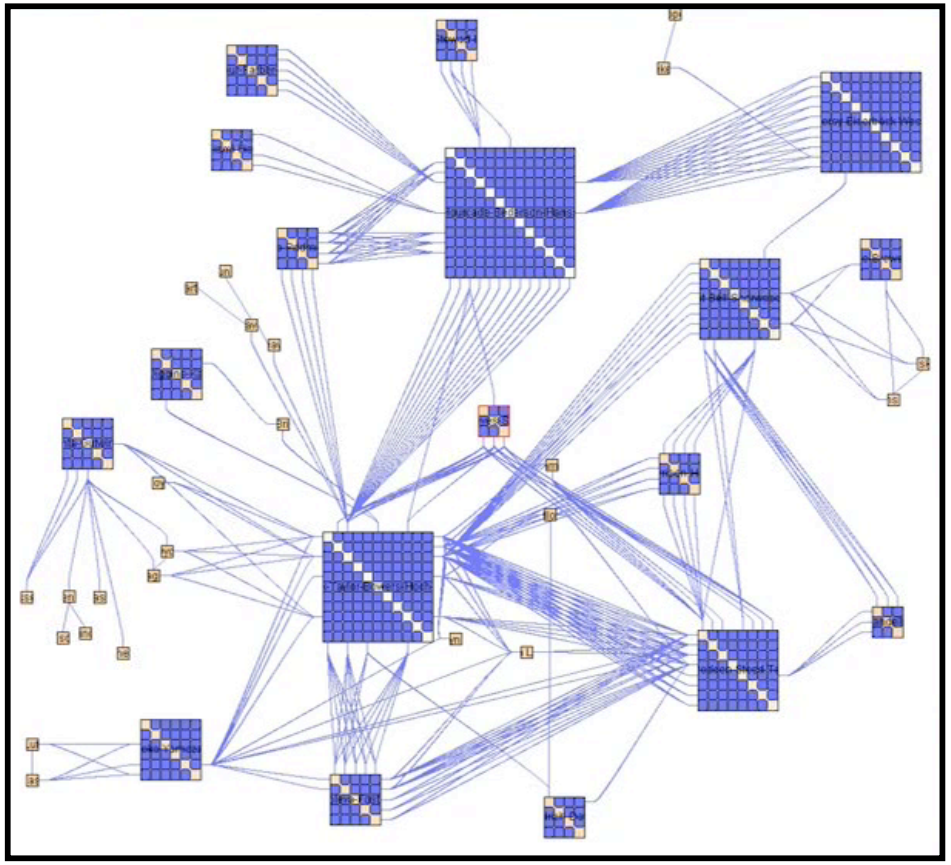
Visual Representation

clone

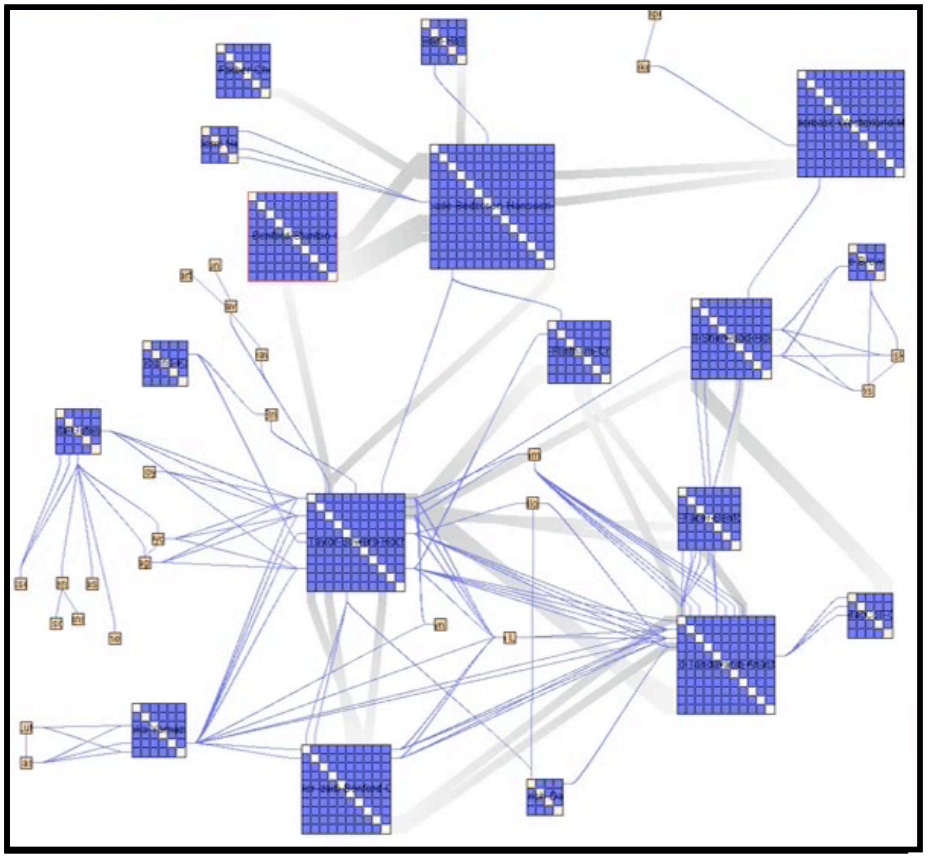
split



Introducing duplicates in a clustered graph



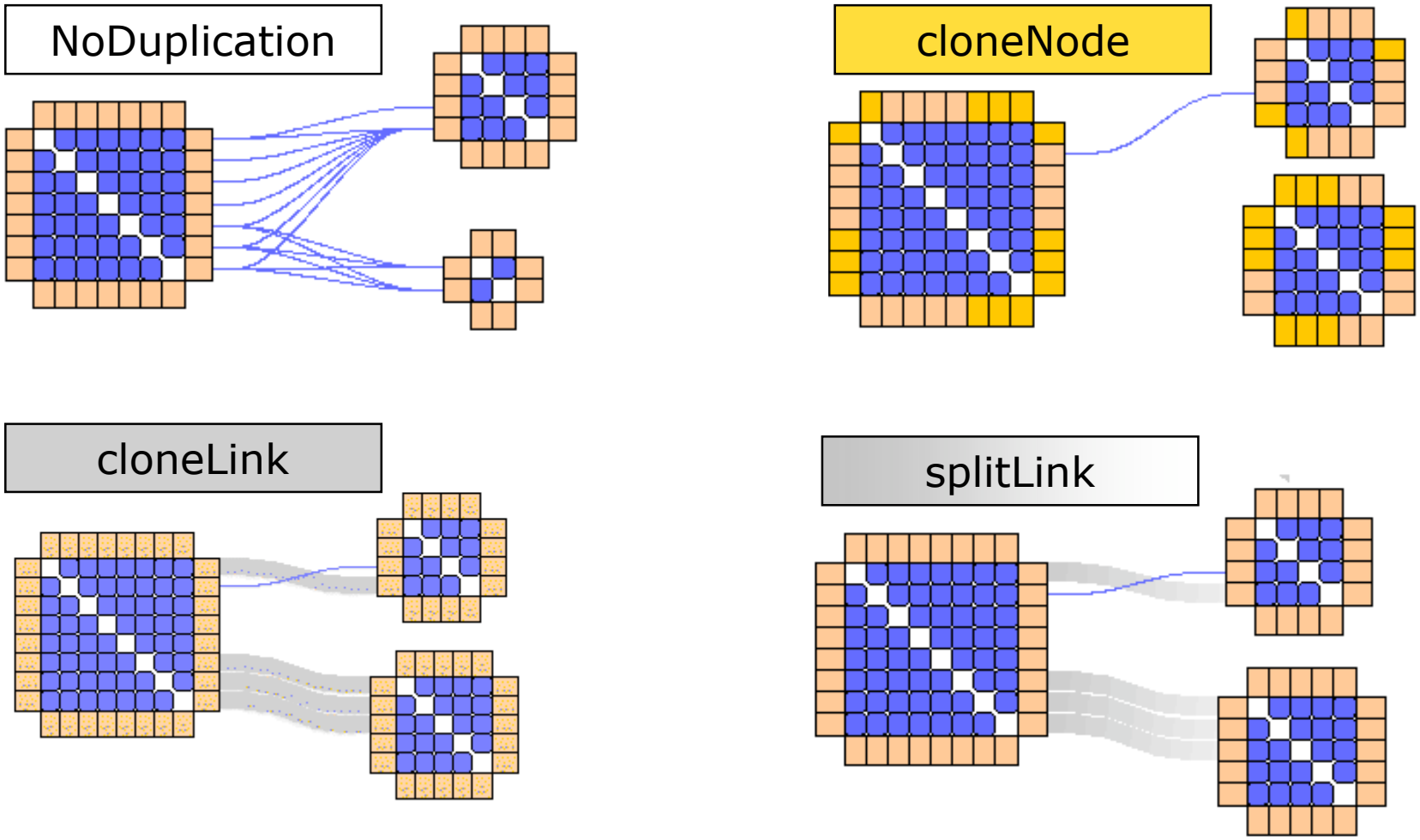
Central actors in one community



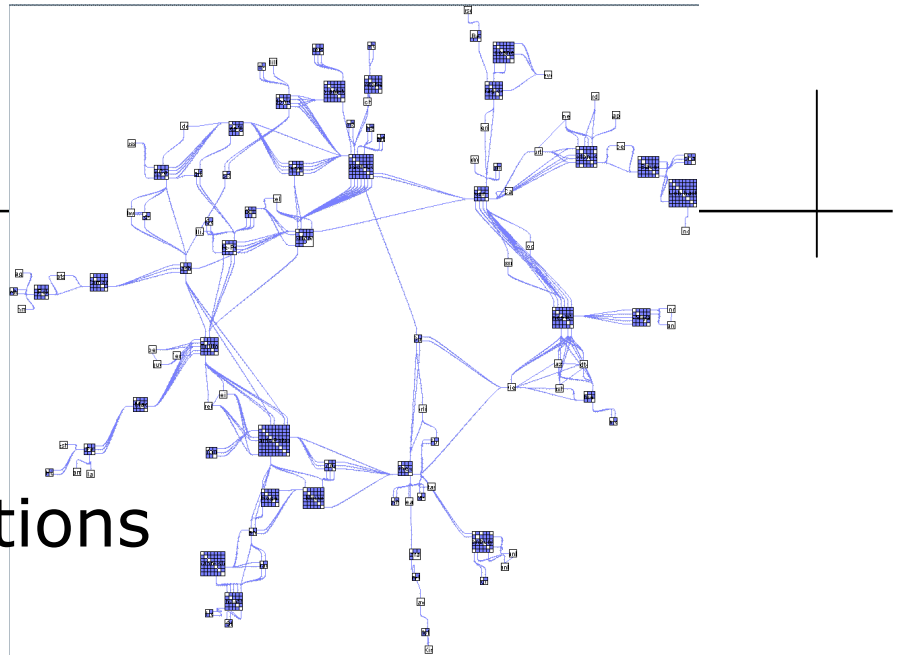
Central actors split-duplicated

Study

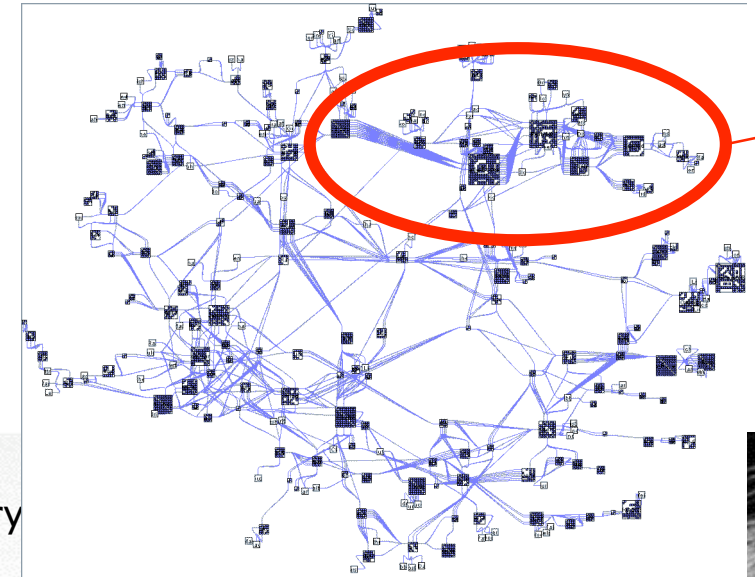
Duplications: an interesting track?

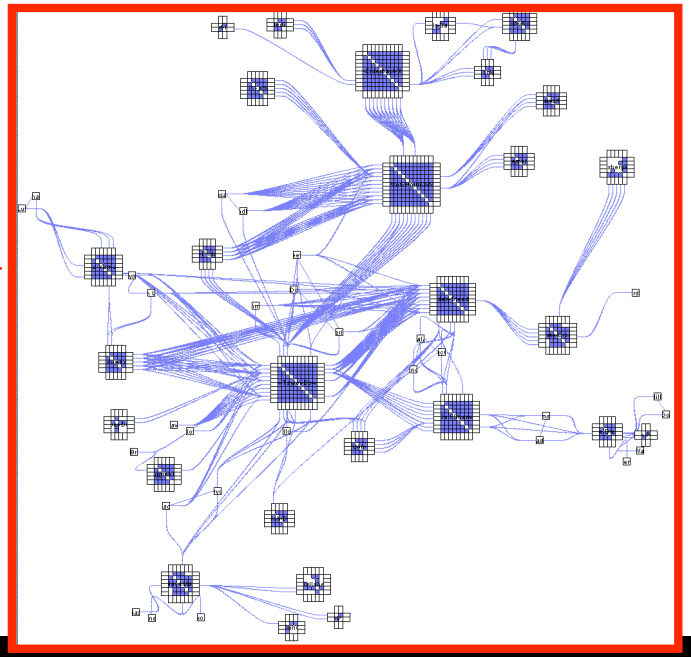


Datasets



- UIST
  - 20 years of publications
- CHI
  - 25 years of publications





Tasks

- Graph readability tasks
  - Number of vertices
  - Shortest distance between a pair of vertices
- Social network analysis tasks

  Community tasks
  - Pair of strongly connected communities
  - Central community

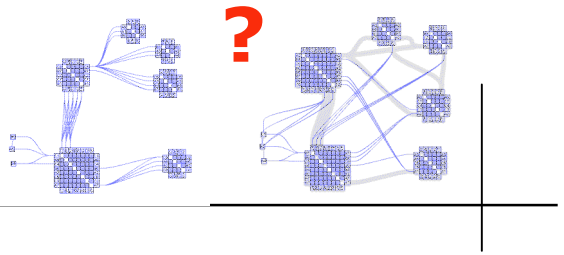
  Important actor tasks
  - Cut point
  - Most connected actor

Procedure

- 12 participants
- Independant variables
  4 visualizations
  x 2 datasets (difficulty)
  x 6 tasks
  x 2 repetitions = 96 trials for each participant
- Dependant variables
  - Success Rate
  - Performance Time
- User Preferrence

Results

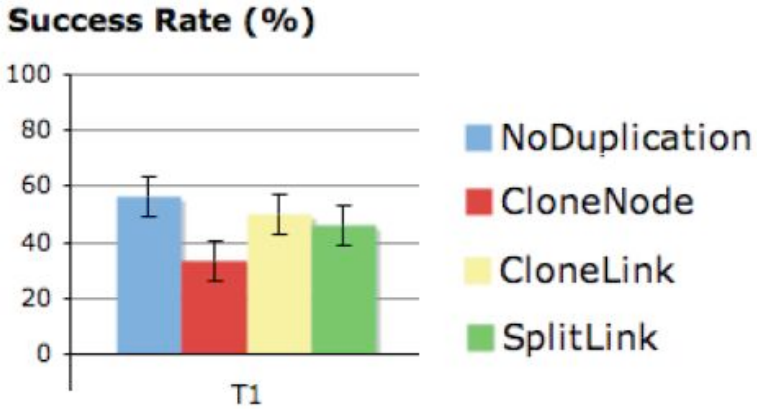
Number of actors



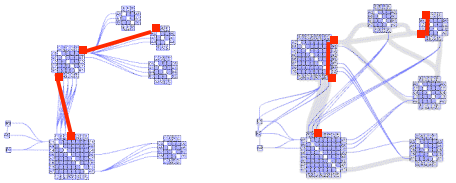
H: Duplication will degrade peformances

- *Surprisingly…*

…not, people are indeed very bad to perform this task with any visualization!



Shortest distance

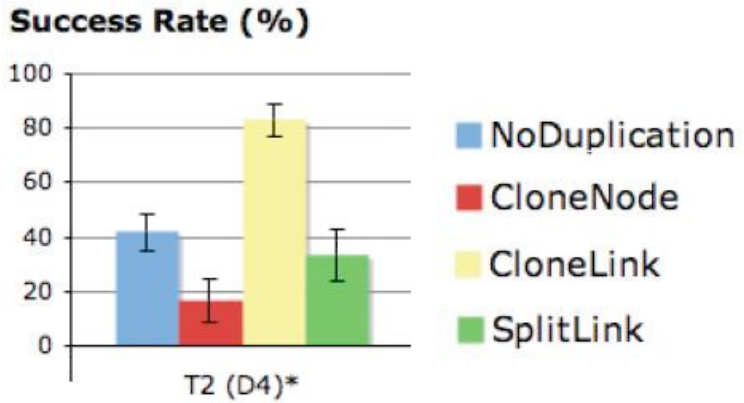


H: Duplications will degrade peformances

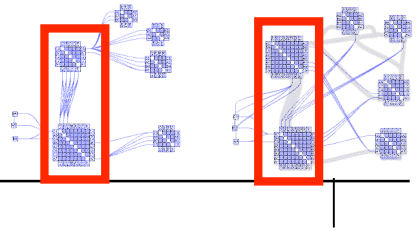
Surprisingly…
No significant difference between SplitLink and NoDuplication
For the difficult case, CloneLink performed even better
→ because clutter is worse than duplications!

A special award to CloneNode for being:

- Error-prone
- Slow
- Frustrating



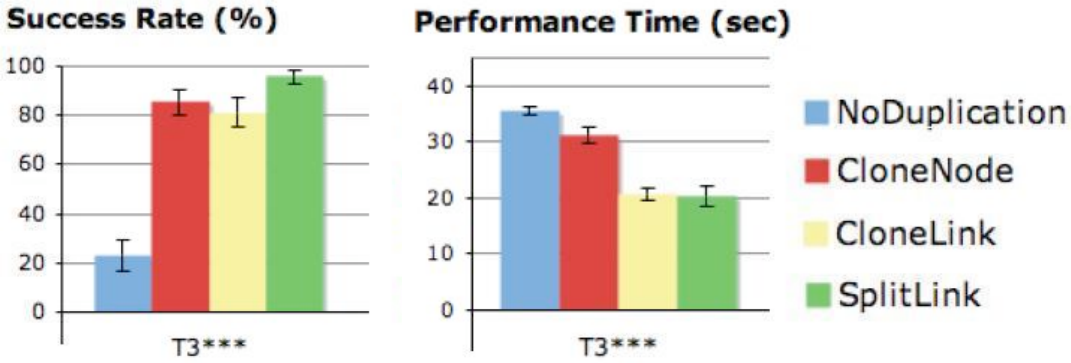
Strongly connected communities



H: Duplications will improve performances, especially the ones with links

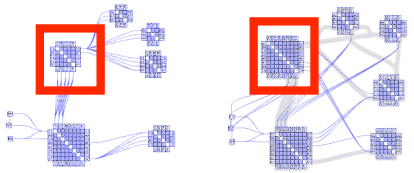
We were right !

Duplications with links were our big winners !



- 11/12 participants preferred the duplications cases

Most central community



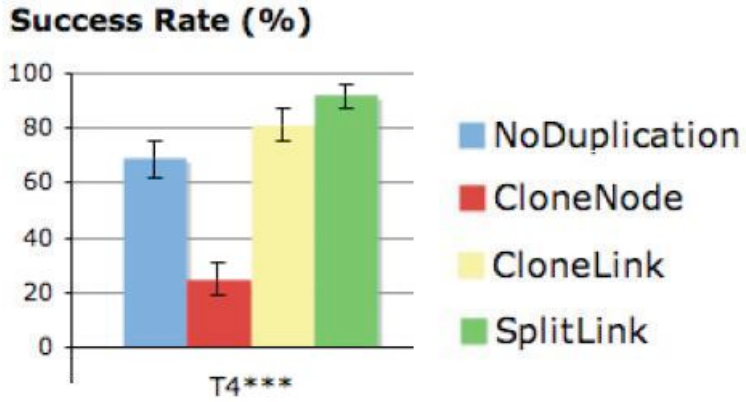
H: Duplications will improve performances

- *Not Surprisingly…*

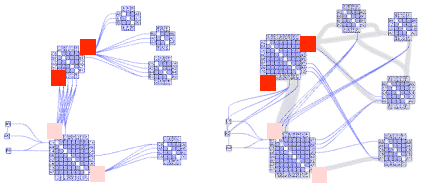
Duplications with Links did improve the success rate, CloneNode did not

- *But Surprisingly…*

no significant difference for the performance time



Most Connected Actor



H: Duplications will make the identification harder

- *Once again…*

  … we were wrong
- Participants were smart on that one

No significant difference between duplication cases and noDuplication

→ most connected actor ≈ in many communities

8/12 participants preferred duplications cases

User feedback

- Several participants *froze* with SplitLink…

  « some missing links? »

- Several participants *froze* with CloneLink…

  « some duplicated links? »

- Almost all participants went *crazy* with CloneNode…

  « where are these x!@kYx! duplicates?! »

User feedback

Improving readability

- « duplication cleans the graph, it makes it easier to see the shortest path »
- « Overall, SplitLink looks **cleaner** but I would need **much more training** to be confident about what I am doing»

Perception of duplication links

- « It is easier to perform tasks using the grey thick lines, because I can ignore them when I want to »
- « [For community tasks,] grey lines tell the story»

Discussion & Conclusion

In summary

- **Graph readability**

Evaluating the number of vertices is hard.
→ is it a NodeTrix artefact?

Shortest path: clutter is worse than duplications

- **Social network analysis**

Community tasks
→ The number of duplicates between communities tells the story

Most connected actor
→ most connected is likely to be in many communities

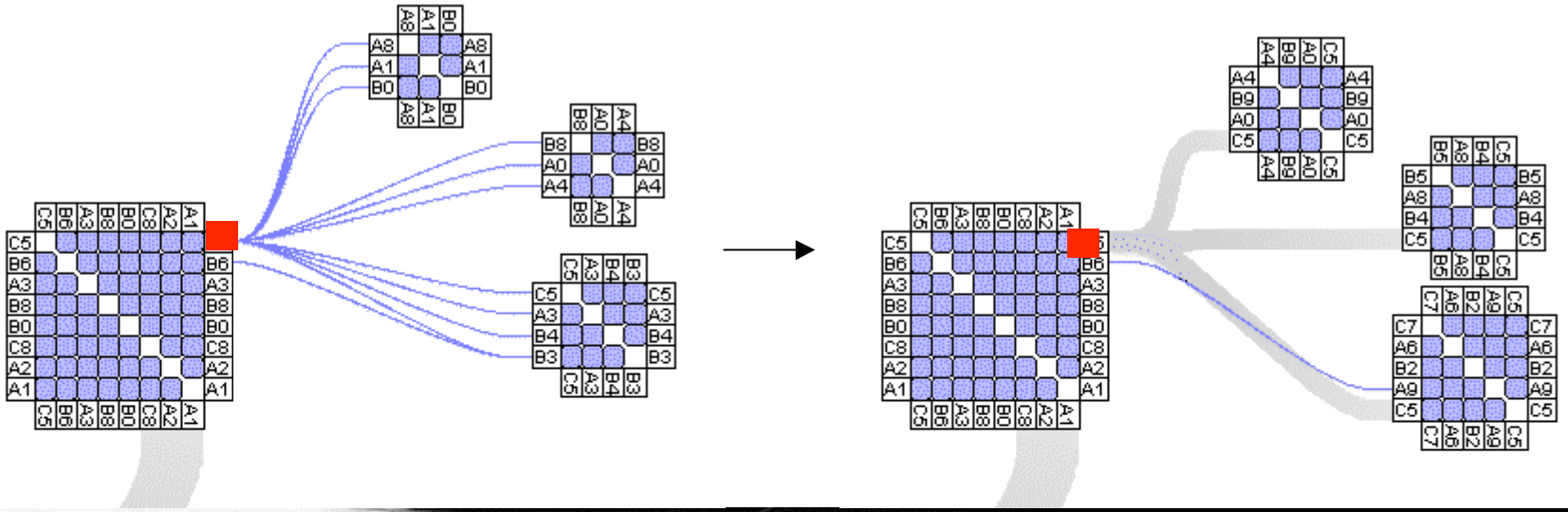
Guidelines

- Why duplicate?
  - To reduce clutter between communities
  - For community tasks
- What type of duplication?
  - Split **OR** Clone
  - Clone as a base case, Split requires training
- How to visualize duplications?
  - Thick links provide **far** better results
  - Highlight is a plus

Next?

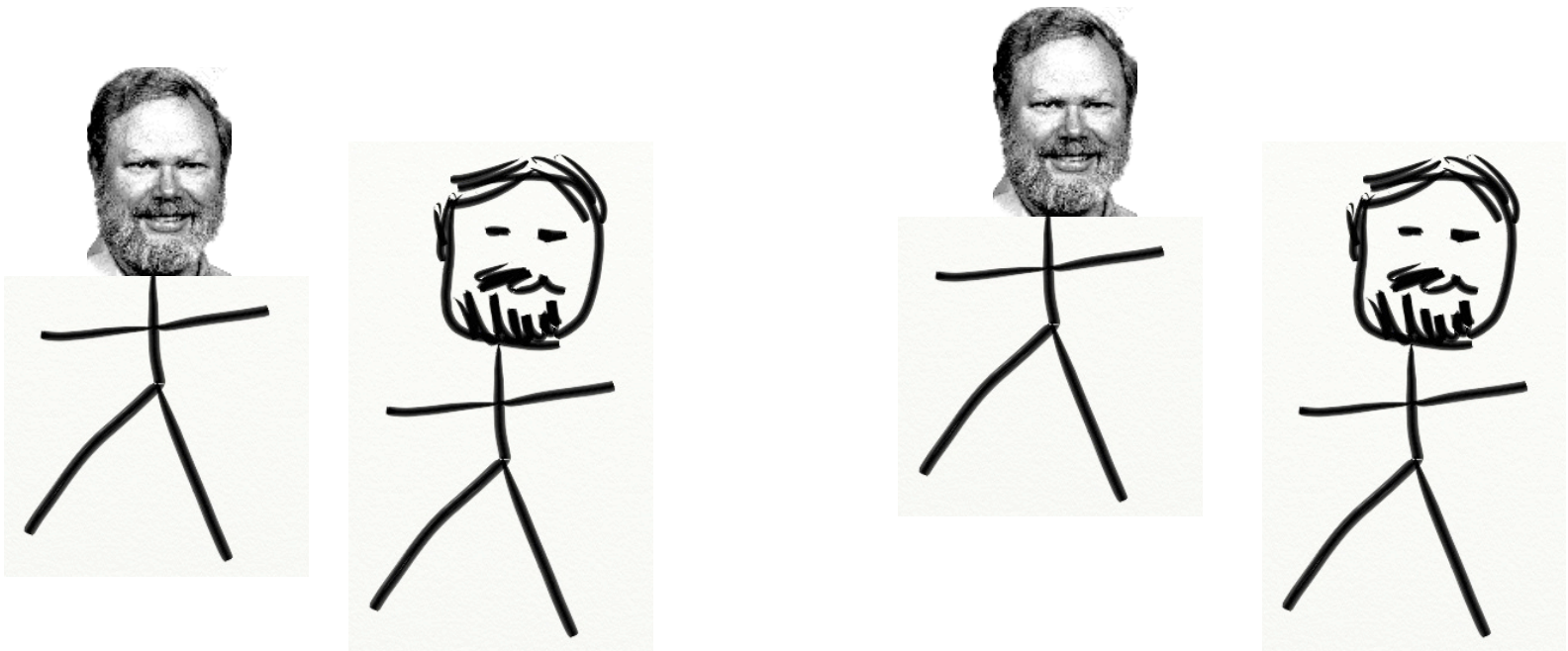
- Interaction
  - To create, edit and visualize duplications

  → interactive merge of duplicates for SplitLink



Thanks for your attention

- I will be happy to answer your questions!



A larger bit of discussion

- Why are results holding with standard node-link diagrams?
  - Intra-community readability improves with NodeTrix but inter-community readability is affected in both cases
  - Duplication affects the inter-community readability